

EXHIBIT B

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

Daily News, LP; The Chicago Tribune Company, LLC; Orlando Sentinel Communications Company, LLC; Sun-Sentinel Company, LLC; San Jose Mercury-News, LLC; DP Media Network, LLC; ORB Publishing, LLC; and Northwest Publications, LLC,

Plaintiffs,

v.

MICROSOFT CORPORATION, OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO, LLC, OPENAI GLOBAL, LLC, OAI CORPORATION, LLC, and OPENAI HOLDINGS, LLC,

Defendants.

Civil Action No. 1:24-cv-03285-SHS

PLAINTIFFS' FIRST SET OF REQUESTS FOR PRODUCTION TO OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, AND OPENAI HOLDINGS, LLC

Pursuant to the Federal Rules of Civil Procedure and the Local Rules of the Southern District of New York, (the “Local Civil Rules”), Plaintiffs Daily News, LP (the “New York Daily News”); The Chicago Tribune Company, LLC, (the “Chicago Tribune”); Orlando Sentinel Communications Company, LLC (the “Orlando Sentinel”); Sun-Sentinel Company, LLC (the “Sun-Sentinel”); San Jose Mercury-News, LLC (the “Mercury News”); DP Media Network, LLC (the “Denver Post”); ORB Publishing, LLC (the “Orange County Register”); and Northwest Publications, LLC (the “Pioneer Press”) (collectively the “Publishers”) request that OpenAI, Inc., OpenAI LP, OpenAI GP, LLC, OpenAI, LLC, OpenAI OPCO LLC, OpenAI Global LLC, OAI Corporation, LLC, and OpenAI Holdings, LLC (“OpenAI Defendants”) produce the requested documents at the offices of Rothwell, Figg, Ernst and Manbeck, P.C., 901 New York Ave NW,

III. DOCUMENT REQUESTS

REQUEST FOR PRODUCTION NO. 1:

Documents concerning the use of each Publisher's Content for training or fine-tuning Your Text Generation AI Models, including, for every Text Generation AI Model that uses any Publisher's Content: (i) the amount and prevalence of each Publisher's Content in the Training Datasets by token count, word count, article count, as a percentage of the Training Datasets, or by any other metric; (ii) rules or guidance for choosing what is and is not included in the Training Datasets for the Text Generation AI Models; (iii) use of each Publisher's Content to clean or curate the Training Datasets; (iv) any weighting of particular Training Datasets for the Text Generation AI Models; and (v) use of each Publisher's Content for any post-training activity, including fine tuning.

REQUEST FOR PRODUCTION NO. 2:

Documents You have gathered to submit or have submitted to any legislative or executive agency, committee, or other governmental entity in the United States or Europe that concern or relate to the allegations in the Complaint, including any Documents related to Your intended or actual submission.

REQUEST FOR PRODUCTION NO. 3:

Documents concerning copyrighted material in the Text Generation AI Models, the Training Datasets, or any Generative AI Products and Services, including documents or communications regarding the appropriateness, value, importance, necessity, and/or concerns over or legal implications of the use of such materials.

REQUEST FOR PRODUCTION NO. 4:

Documents concerning claims of copyright and trademark infringement, including complaints regarding infringement, steps taken to respond to such claims, responses to such complaints, steps taken to mitigate any harm resulting from such claims, and steps taken to prevent infringement.

REQUEST FOR PRODUCTION NO. 5:

Documents reflecting policies, procedures or practices concerning use of intellectual property in AI Models, including the policy OpenAI referred to as “Copyright Shield” in the November 6, 2023, blog post on Your website titled “New models and developer products announced at DevDay.”

REQUEST FOR PRODUCTION NO. 6:

Documents concerning alternatives to using copyrighted content to train the Text Generation AI Models without compensation including using licensing agreements or material in the public domain.

REQUEST FOR PRODUCTION NO. 7:

Documents concerning the impact of AI Models on Journalism and the past, current or future use of the Text Generation AI Models or any Generative AI Products and Services as a source of Journalism, substitute for Journalism, or product in competition with Journalism (including any of the Publisher’s Journalism).

REQUEST FOR PRODUCTION NO. 8:

Documents sufficient to show to what extent and in what form each Publisher’s Content resides in, is stored in, is copied by, or otherwise exists within the Training Datasets and/or the Text Generation AI Models.

REQUEST FOR PRODUCTION NO. 9:

Documents sufficient to show and describe OpenAI’s corporate organizational structure,

including those sufficient to show OpenAI's board of directors, parent companies, subsidiaries, affiliates, departments, and divisions; and those sufficient to show the function each Defendant plays within Your corporate structure as a whole.

REQUEST FOR PRODUCTION NO. 10:

Documents sufficient to identify OpenAI's current and former officers, directors, managers, employees and consultants with knowledge of the Text Generation AI Models, business practices regarding the Text Generation AI Models, and commercialization and use of the Text Generation AI Models.

REQUEST FOR PRODUCTION NO. 11:

Documents concerning OpenAI's transition to a for-profit company, as alleged in paragraphs 53-54 of the Complaint.

REQUEST FOR PRODUCTION NO. 12:

Documents concerning Defendants' use of Retrieval Augmented Generation ("RAG") in Generative AI Products and Services, including: (i) use of each of the Publisher's Content for RAG; (ii) rules or guidance for preferring or avoiding certain web domains (including each of the Publishers' web domains or sites known to host content copied from other sources) when choosing what to include in a generative result; (iii) rules for whether and how the results from RAG search products are similar to or different from standard search results; (iv) and rules regarding paywalls.

REQUEST FOR PRODUCTION NO. 13:

Documents concerning the features or capabilities of Retrieval Augmented Generation.

REQUEST FOR PRODUCTION NO. 14:

A complete listing of every data sample in the Training Datasets used to train each of the Text Generation AI Models and the source from which each data sample was obtained.

Services, including: (i) the ordering of training samples containing the Publishers' or Journalism Content in the training process as compared to training samples not containing the Publishers' or Journalism Content; (ii) placement of training samples containing the Publishers' or Journalism Content in the training process; (iii) rules for choosing the order of training samples containing the Publishers' or Journalism Content in the training process; (iv) rules and procedures for training samples containing the Publishers' or Journalism Content that are not used on all other training samples; (v) the processing of the Publishers' or Journalism Content in the fine-tuning/instruction tuning phase of training; (vi) use of the Publishers' or Journalism Content to train or inform certain, modules, layers, mixtures of experts, or other portions of the Text Generation AI Models more than others, (vii) any measures taken to include or exclude certain portions of training samples containing the Publishers' or Journalism Content during training, and (viii) the importance of training on accurate and/or well-written content, including the Publishers' or Journalism Content.

REQUEST FOR PRODUCTION NO. 20:

Documents sufficient to show the total number of records and tokens in Defendants' Training Datasets and the total number of records and tokens that contain (i) the Publishers' Content, (ii) Journalism Content, and (iii) copyrighted content.

REQUEST FOR PRODUCTION NO. 21:

Documents sufficient to show each source of records in Defendants' Training Datasets, and for each source: (i) the total number of tokens, (ii) the total number of records, (iii) the total number of tokens that contain the Publishers' Content, (iv) the total number of records that contain the Publishers' Content, (v) the total number of tokens that contain Journalism content, (vi) the total number of records that contain Journalism content; (vii) the total number of tokens that contain copyrighted content, (viii) the total number of records that contain copyrighted content and (ix) any agreements relating to your use, licensing, or purchasing of the source of records.

REQUEST FOR PRODUCTION NO. 22:

Documents sufficient to identify each article of the Publishers' Content contained in Defendants' Training Datasets.

REQUEST FOR PRODUCTION NO. 23:

Documents concerning all steps taken or contemplated to train, instruct, program, or refine Defendants' Generative AI Products and Services using the Publishers' Content or Journalism Content, including: (i) the use of the Publishers' or Journalism Content in instruction/fine tuning; (ii) use of the Publishers' or Journalism Content in Reinforcement Learning from Human Feedback ("RLHF") or Reinforcement Learning from AI Feedback ("RLAIF"), (iii) use of the Publishers' or Journalism Content to train the reward component or function of RLHF or RLAIF, (iv) use of the Publishers' or Journalism Content to evaluate a feature of a Training Dataset, such as classifying Journalism Content or identifying the quality of a training sample, and (v) RLHF moderator and ranking guidance, training materials and ranking factors, weighting or scoring factors, and results/output of the human annotators, moderators and reviewers.

REQUEST FOR PRODUCTION NO. 24:

Documents concerning any analyses, studies, measurements, testing, experimentation, assessments, or other evaluation of the incremental value of including Journalism Content, including the Publishers' Content, in Defendants' Training Datasets.

REQUEST FOR PRODUCTION NO. 25:

Documents concerning any analyses, studies, measurements, testing, experimentation, assessments, or other evaluation of the quantity, ordering, placement, or other configurable parameter of Journalism Content, including the Publishers' Content, in Defendants' Training Datasets.

REQUEST FOR PRODUCTION NO. 26:

Documents concerning Your use or knowledge of web crawlers, bots, spiders, user agents, and related tools to access the Publishers' Content, including the Publishers' websites and digital

the United States, including price quotes, price lists, subscription amounts, price announcements, and actual prices paid by customers.

REQUEST FOR PRODUCTION NO. 74:

All documents that you intend to rely upon at trial, in support of any motion, or to present to the Court for any purpose in this litigation.

REQUEST FOR PRODUCTION NO. 75:

All documents produced or otherwise provided by You to The New York Times Company in the litigation related to this case.

Dated: June 14, 2024

/s/Steven Lieberman

Steven Lieberman (SL8687)
Jennifer B. Maisel (5096995)
Robert Parker (*admitted pro hac vice*)
Jenny L. Colgate (*admitted pro hac vice*)
Mark Rawls (*admitted pro hac vice*)
Kristen J. Logan (*admitted pro hac vice*)
Bryan B. Thompson (6004147)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
901 New York Avenue, N.W., Suite 900 East
Washington, DC 20001
Telephone: (202) 783-6040
Facsimile: (202) 783-6031
slieberman@rothwellfigg.com
jmaisel@rothwellfigg.com
rparker@rothwellfigg.com
jcolgate@rothwellfigg.com
mrawls@rothwellfigg.com
klogan@rothwellfigg.com
bthompson@rothwellfigg.com

Jeffrey A. Lindenbaum (JL1971)
ROTHWELL, FIGG, ERNST & MANBECK, P.C.
3 Manhattanville Road, Suite 105
Purchase, New York 10577

Telephone: (202) 783-6040
Facsimile: (202) 783-6031
jlindenbaum@rothwellfigg.com

Attorneys for Plaintiffs

CERTIFICATE OF SERVICE

It is hereby certified that, on June 14, 2004, the foregoing **PLAINTIFFS' FIRST SET OF REQUESTS FOR PRODUCTION TO OPENAI, INC., OPENAI LP, OPENAI GP, LLC, OPENAI, LLC, OPENAI OPCO LLC, OPENAI GLOBAL LLC, OAI CORPORATION, LLC, AND OPENAI HOLDINGS, LLC** was served by email upon the following counsel of record for OpenAI Defendants:

<p>Latham & Watkins LLP 1271 Avenue of the Americas New York, NY 10020 212-906-1200 Allison Levine Stillman Email: alli.stillman@lw.com</p> <p>Latham & Watkins LLP 555 Eleventh Street NW Suite 1000 Washington, DC 20004 202-637-2200 Sarang Damle Email: sy.damle@lw.com Elana Nightingale Dawson Email: elana.nightingaledawson@lw.com</p> <p>Latham & Watkins, LLP 505 Montgomery Street, Ste. 2000 San Francisco, CA 94111 (415)-391-0600 Andrew Gass Email: andrew.gass@lw.com Joseph Richard Wetzel, Jr. Email: joe.wetzel@lw.com</p>	<p>Morrison & Foerster LLP 425 Market Street San Francisco, CA 94105 415-268-7000 Joseph C. Gratz Email: JGratz@mofo.com Vera Ranieri Email: vranieri@mofo.com</p> <p>Morrison & Foerster LLP 707 Wilshire Boulevard Suite 6000 Los Angeles, CA 90017-3543 213-892-5200 Allyson R. Bennett Email: ABennett@mofo.com</p> <p>Morrison & Foerster LLP 250 West 55th Street New York, NY 10019-9601 212-468-8000 Eric K. Nikolaides Email: ENikolaides@mofo.com</p> <p>KEKER, VAN NEST & PETERS LLP 633 Battery Street San Francisco, CA 94111 415 391-5400 Robert A. Van Nest Jr Email: rvannest@keker.com Paven Malhotra Email: pmalhotra@keker.com R. James Slaughter Email: rvannest@keker.com Nicholas S. Goldberg Email: ngoldberg@keker.com</p>
---	---

	Michelle Ybarra Email: mybarra@keker.com Katie Lynn Joyce Email: kjoyce@keker.com Thomas E. Gorman Email: tgorman@keker.com
--	---

/s/Nasri V. B. Hage
Nasri V. B. Hage